DOCUMENT RESUME

ED 378 219 TM 022 585

AUTHOR Baker, Eva L.; And Others

TITLE The Apple Classrooms of Tomorrow. The UCLA Evaluation Studies.

INSTITUTION California Univ., Los Angeles. Center for Technology Assessment.; National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.

SPONS AGENCY Apple Computer, Inc., Cupertino, CA.
REPORT NO CSE-TR-353
PUB DATE 93
NOTE 24p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Access to Education; *Computer Uses in Education; *Educational Technology; Elementary Education; Interaction; Measurement Techniques; *Outcomes of Education; Program Evaluation; Research Design; *Teacher Attitudes; Teaching Methods; *Triangulation
IDENTIFIERS *Apple Classrooms of Tomorrow; University of California Los Angeles

ABSTRACT

The Apple Classrooms of Tomorrow (ACOT) project was initiated in classrooms at five sites as a program of research on the impact of interactive technologies on teaching and learning. ACOT provided teachers and students with an Apple computer at home and at school. The University of California, Los Angeles, conducted an evaluation of the ACOT program using a triangulation approach, with measurement tools developed for the study. This report summarizes evaluation findings from 1987 through 1990. Evaluation studies documented the challenges of assessing the effects of technology access on teaching and learning. While not conclusive, the study indicates that ACOT may have had selective positive impact on writing and on student attitudes, particularly as projects matured. These results make evident the need to examine the impact of specific instructional uses of technology on specific outcomes. Although there were findings that changes in practice, such as increased cooperative learning and decreased "directive" teaching, were consistent with ACOT goals, these findings do not reveal the particular role of technology in these changes. These issues affirm the importance of measures sensitive to particular technology project designs and contexts. (Contains 18 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

CRESST National Center for Research on Evaluation, Standards, and Student Testing

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it
☐ Minor changes have been made to improve reproduction quality
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
J. C. BEER
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED 378 219

The Apple Classrooms of Tomorrow℠
The UCLA Evaluation Studies

CSE Technical Report 353

Eva L. Baker, Maryl Gearhart, and Joan L. Herman

Center for the Study of Evaluation/
Center for Technology Assessment
University of California, Los Angeles

- UCLA Center for the Study of Evaluation

in collaboration with:

- University of Colorado
- NORC, University of Chicago
- LRDC, University of Pittsburgh
- The RAND Corporation

TM022585

BEST COPY AVAILABLE

The Apple Classrooms of Tomorrow[sm]
The UCLA Evaluation Studies

CSE Technical Report 353

Eva L. Baker, Maryl Gearhart, and Joan L. Herman

Center for the Study of Evaluation/
Center for Technology Assessment
University of California, Los Angeles

January 1993

Center the Study of Evaluation
Graduate School of Education
University of California, Los Angeles
Los Angeles, CA 90024-1522
(310) 206-1532

Copyright © 1993 The Regents of the University of California

The work reported herein was supported by the Advanced Development Group of Apple Computer, Inc. The findings and opinions expressed in this report are solely those of the authors.

THE APPLE CLASSROOMS OF TOMORROW[sm]

THE UCLA EVALUATION STUDIES[1]

Eva L. Baker, Maryl Gearhart, and Joan L. Herman

Center for the Study of Evaluation/Center for Technology Assessment

University of California, Los Angeles

Background

The Apple Classrooms of Tomorrow[sm] (ACOT[sm]) project was initiated in classrooms at five school sites in 1985 as a program of research on the impact of interactive technologies on teaching and learning. Originally conceived as a program to study what happens when "tomorrow's" resources are routinely available in classrooms, ACOT provided students and teachers an Apple computer both at school and at home. Sites were selected by ACOT staff to represent a range of student, school, and community characteristics.[2] The process of site selection differed across sites, although all sites and their participating teachers were required to demonstrate their interest and willingness to participate.

While the project has expanded over time to encompass a larger and more diverse set of efforts, key components at all sites have been the provision of high technology access, site freedom to develop technology-supported curriculum and pedagogy as appropriate to site goals, and the resulting study of what happens when technology support is readily available to students and

---

[1] We are grateful for the cooperation of ACOT site coordinators and the teachers, principals, parents, and students at ACOT and comparison schools. Warm thanks to our talented research staff: John Novak, Andrea Whittaker, David Niemi, Howard Herl, Darlene Galluzzo, and Patricia Mutch.

[2] Elementary sites are located in northern California's Silicon Valley, a suburban Tennessee community, an urban Tennessee setting, and a rural Minnesota location. The secondary site is located within a major city in Ohio.

teachers. ACOT has encouraged instructional innovation, emphasizing to participating teachers the potential of computers to support student initiative, long-term projects, access to multiple resources, cooperative learning, and instructional guidance rather than stand-up teaching.

From 1987 through 1990, UCLA conducted a series of evaluation studies focused on the five original ACOT sites (Baker, Gearhart, & Herman, 1990, 1992; Baker, Herman, & Gearhart, 1988, 1989). Assessment of ACOT presented a continual challenge to what is meant by "formative evaluation," requiring ongoing attention to new goals, new modes of instructional transaction, and new outcomes as they evolved. ACOT's evolutionary character required both close-up interaction with sites to understand the changes that were occurring and new tools that could capture these changes and their outcomes.

Our approach to the study of ACOT effectiveness was one of triangulation. Recognizing the imperfections of existing measures and the constraints on the real world laboratories in which ACOT was implemented, we employed a strategy to assess progress based upon a range of measures and multiple benchmarks. Comparisons of ACOT students' basic skills performance to nationally reported norms was one approach; comparisons of student progress and achievement over time was another; comparisons of ACOT classrooms with demographically similar classrooms was still another information point; gathering data on classroom practices and parents' background characteristics to help explain student outcomes was yet another. Our strategy was inherently developmental. We started with standard measures and then developed an expanded set of measurement tools to capture ACOT as it evolved. Creating alternative indices of classroom process and student outcomes was a key component in our assessment strategy.

Thus, key attributes of the UCLA evaluation were:

- collection and analysis of a broad range of potential student outcomes;
- collection and analysis of such information over time;
- linking outcome data with information on instructional process to provide explanatory power for findings;
- linking multiple indicators of key outcomes to strengthen the validity of findings;

- combining the strengths of both quantitative and qualitative methodologies;
- using the known characteristics of existing measures as a means of developing and validating new measures;
- providing uniform data collection strategies and measures across the diverse ACOT sites, but reserving places for interests, measures, and effects unique to each site.

Four basic questions guided the work:

1. What is the impact of ACOT on students?
2. What is the impact of ACOT on teachers' practices and classroom processes?
3. What is the impact of ACOT on teachers professionally and personally?
4. What is the impact of ACOT on parents and home life?

This report summarizes our findings from 1987 through 1990. As we will show, our inferences have been challenged by the concurrent evolution of ACOT goals, the need for development of new measures, the resistance of comparison classrooms to participation, and year-to-year changes in project organization at some of the sites.

Methods

Overview

Data collection was initiated in the spring of 1988 with the administration of a range of existing student outcome measures whose characteristics were well-known. Student outcomes assessed in the initial year included: (a) achievement on standardized tests, (b) performance in written composition, and (c) student attitudes. Results were used to provide a baseline for comparison with subsequent administrations in 1989 and 1990. Results also provided a basis for evaluating the usefulness of the information these kinds of measures can produce, and various of the student measures contained in the initial baseline battery were later replicated, revised, replaced, or supplemented as we gained understanding of ACOT outcomes and their appropriate measurement.

Comparison groups were identified by the fall of 1988 for most classrooms at most sites, but because the criteria for their selection varied as a function of site-specific constraints and because the participation of these classrooms was generally problematic, interpretation of comparison data required utmost caution.

To provide evidence of ACOT impact on teachers and instruction and on parents and home activities, teacher and parent surveys were administered in the spring of both 1989 and 1990. Using a quantitative classroom observation method developed in 1988-89, classrooms were observed at two sites in 1989-90 (and in 1990-91, as a follow-up to our core evaluation). Both the survey and observation data provided an explanatory context for the student outcome results.

Student Achievement

Iowa Tests. The Iowa Tests of Basic Skills were selected as the norm-referenced tests because, compared with other norm-referenced instruments, they allowed us to compare students' performances across the many grade levels of ACOT. At the elementary level, we focused on four subtests which reflected core emphases at most sites: Vocabulary, Reading Comprehension, Mathematics Concepts, and Work-Study Skills/Visual Materials. At the secondary level, we chose two subtests, Vocabulary and Social Studies. Four sites agreed to administer the Iowa tests, although at some sites only selected ACOT grade levels completed them, limiting our potential for analysis. Beginning with the spring of 1988, tests were administered in the fall (incoming students) and spring to ACOT and comparison students.

Writing. Students responded to writing prompts that asked for either narrative, descriptive, persuasive, or expository writing. In 1988 and 1988-89, students provided samples of three genres at each site. In 1990, the sites specified the genres most emphasized at each grade level. The writing prompts were derived from those used in the International Association for the Study of Educational Achievement (IEA) Study of Written Composition (Baker, 1987).

Because the IEA study included national samples of students in Grades 6 and 10, ACOT students' essays in Grades 5 and up were rated on scales employed in the IEA study, and results were compared with IEA national

samples. Beginning in the fall of 1988, the essays of the youngest ACOT students (Grades 2-4) were rated with a rubric developed by a southern California district for third- and fifth-grade narrative writing; the scales were derived from the same sources as the IEA scales. Results for ACOT students were compared with results from students from this southern California district.

Student Attitudes

Five approaches were used to assess student attitudes. With exceptions noted below, the self-report measures were administered on the same schedule as the Iowa tests (see above).

1. **Normed measures**: *School Attitude Measure (SAM)* is a survey instrument published by Scott-Foresman consisting of five scales: Motivation for Schooling; Academic Self-Concept, Performance-Based; Academic Self-Concept, Reference-Based; Sense of Control; and Instructional Mastery. *SAM* is normed for Grades 4-12 and was administered to grades within this range. *Self-Concept and Motivation Inventory (SCAMIN)* is a self-report survey instrument published by Person-O-Metrics, Inc. (Dearborn Heights, MI) consisting of four scales at the early elementary level (Achievement Needs, Achievement Investment, Role Expectations, and Self-Adequacy). It is normed for K-12 and was administered to grades below fourth.

2. ***UCLA Student Questionnaire*** (spring 1990 only): Closed- and open-ended item content addressed a range of possible affective outcomes, some containing specific reference to technology impact, and others not.

3. **Student attendance and mobility patterns** were examined for sites that were able to provide us these data.

4. **Teacher questionnaire and interview** items were analyzed for teachers' perceptions of students' attitudes.

5. **Parent questionnaire and interview** items were examined for parents' perceptions of students' attitudes.

Teachers' Practices, Perceptions, and Attitudes

Questionnaires. UCLA-designed surveys were administered to ACOT teachers in the spring of both 1989 and 1990 and adapted for administration to comparison teachers. The 1989 questionnaire focused on curricular practices, perceptions of students' achievement, and perceived professional growth and

stress. In addition, teachers were asked to complete the *Occupational Stress Inventory* (Psychological Assessment Resources, Lutz, FL). The 1990 Teacher Questionnaire focused on changes in curriculum and teaching practices, sources of influence on educational practice, perceptions of students' achievement, and perceived stress. The rate of return varied markedly by site and group.

Classroom observations of instructional practice. The observation scheme was designed in 1988 to capture the nature of instructional tasks, the roles of teachers, the nature of social relationships, the variety of resources, and the responses of students (Gearhart, Herman, Baker, Novak, & Whittaker, in press). Major categories of the scheme include: subject area of instruction, classroom organization, teachers' instructional roles, symbol systems in use, length of students' responses, the level of challenge of students' work, the variety of resources in use, and students' responses to the activities (appropriateness of behavior, focus and investment, productivity of peer interaction). The observations are recorded in timed intervals (5 minutes for elementary level, 10 minutes for secondary level) and supplemented with field notes.

Impact on Parents and the Home

UCLA-designed surveys were administered to ACOT parents in the spring of both 1989 and 1990 and adapted for administration to comparison parents. Both closed- and open-ended items addressed parents' perceptions of the impact of ACOT on their children, parental aspirations, and uses of the home computer. The 1990 questionnaire was similar in content, although certain items were revised to clarify their intent. The rate of return varied considerably by site and group, and response by comparison parents was particularly weak. (In 1990, only two sites agreed to distribute the questionnaires to comparison parents.)

Summary of Findings

Firm conclusions about the effects of ACOT on student and other outcomes are constrained by a number of factors:

- ACOT was implemented in a relatively small number of classrooms, often only one at a particular grade level, dispersed over an

intentionally diverse set of school sites; thus ACOT effects were confounded by district requirements, teachers' influences, curriculum selection, school ambiance, and the characteristics of students, among other factors.

- Identification and participation of comparison classrooms were problematic.

- There were occasional site requests for adaptations of our instrumentation (particularly for the 1989-90 year), adaptations that challenged meaningful comparisons with national samples and interpretation of longitudinal comparisons.

- Finally, at some sites there were changes made from year to year in classroom and/or grade level ACOT participation which rendered longitudinal comparisons inappropriate.

Nonetheless, we felt that some inferences were possible.

Student Outcomes

In general, the results suggested that ACOT students had at least maintained their performance levels on standard measures of educational achievement in basic skills and had sustained positive attitudes. We drew these conclusions from measures addressing the traditional activities of schooling: standardized achievement tests, writing assessments, normed student attitude instruments, and archival data on student attendance. Generally, the ACOT program appeared as effective in promoting commonly measured student outcomes as the more typical instructional programs provided by the comparison sites.

There were several variations on this overall theme:

- Where we had comparison results, there was evidence of *domain-specific ACOT impact*. There were indications that ACOT enhanced students' writing abilities at one site in 1988-89, and three in 1989-90 (for a total of three sites across the two years). Impact in other domains of student achievement was negligible or nonexistent.

- There was also evidence of *site-specific impact*: At one of three sites with comparison students, ACOT students' performance on achievement tests was higher on some subtests; at one site ACOT students' performance was lower than that of comparison students for some subtests.

- There was also evidence that *project maturity* may have mediated more consistent cross-site ACOT impact on students' attitudes.

Compared with the results of 1988-89, the 1989-90 results for student attitudes were more positive and more consistent across sites. In 1990 the attitudes of most ACOT children were at or above the national median for our normed measures, and at one of our two sites with comparison students (for these measures), there was evidence of improvement over time for ACOT students only.

We viewed the overall maintenance of the *status quo* positively. First, the ACOT environment required both teachers and students to adapt to a host of new technological options. Student time spent in learning word processing and other software was time that otherwise would have been spent on traditional schools subjects. Similarly, the time teachers needed to acquire technology skills and familiarity with software appropriate to individual and/or grade level needs or particular curricular objectives might have resulted in less time spent in curriculum planning. Time too may have been lost to occasional technical failures. Any of these short-term problems could have resulted in less-than expected student academic growth or a temporary undermining of students' motivation and attitudes. Yet negative consequences on student outcomes were not in evidence.

Second, the ACOT experience appeared to be resulting in significant new learning experiences requiring higher-level reasoning and problem solving (see **Instructional impact** below). Because more time in long-term, constructive activities may translate into less time in basic skills instruction, some decrement in basic skills test performance also might have been expected. From both these vantage points, then, maintenance of expected performance levels could be viewed as an accomplishment.

Effects on Teachers and Instruction

Our goal was to examine how characteristics of teachers and contexts interacted with ACOT's impact on teachers and instruction. From 1988 to 1989, our investigation relied on teacher questionnaires supplemented with field notes from site visits. Revised in 1989-90, our second questionnaire focused on changes in practices, and an observation instrument was developed to document classroom activities in two sites. Results from all measures indicated that the ACOT experience affected instruction and impacted teachers personally and professionally, but it was not possible to distinguish clearly which factors were responsible for which changes. At each site, ACOT represented a different pattern of opportunities for technology access,

professional development, and restructuring of collegial roles and relationships. These factors—in combination with variation in survey returns and the virtual nonparticipation of comparison teachers—made inferences difficult. The results are therefore reported as "broad brush" portraits of impact.

Instructional impact. Based on 1989 and 1990 surveys, ACOT teachers' reports of their classroom practices indicated fairly typical subject matter emphases and pedagogical methods, along with occasional experimentation with the use of technology to support innovative practices such as cooperative group work or long-term student projects. In 1990, teachers' self-descriptions of instructional change were consistent with many changes encouraged by ACOT: Teachers reported increases in classroom organizations and teaching methods that support student initiative and independence, decreases in traditional stand-up teaching and reliance upon published teachers' guides, and changes in their uses of technology toward less frequent use of published instructional software and more frequent use of applications as well as computer-supported activities of their own design. There was evidence of influence other than ACOT as well, however: For instruction in core subject areas, there were site differences in use of state and district instructional guidelines that appeared to reflect variation in local pressures on student achievement. In addition, teachers acknowledged an impact of both computer access and national interest in writing instruction: Teachers reported teaching students more types of writing and a writing process approach.

Our quantitative observations provided us a direct view of technology's functions in instructional practice, and results suggested effects mediated by teachers' prior practices and expertise (Gearhart, Herman, Baker, Novak, & Whittaker, in press; Gearhart, Herman, Whittaker, & Novak, 1991). In one set of analyses, instructional patterns at an elementary ACOT site differed both by subject area and by the nature of computer use (e.g., use of instructional software vs. use of applications such as word processing). Our data suggested that the ACOT teachers were making motivated choices about resources and pedagogical methods based on subject area, and thus their classroom activities were not "technology-driven" in any simple way. In a second set of analyses, we found an association between school level and degree of instructional innovation: Compared with secondary ACOT teachers, the elementary level

ACOT teachers in our samples appeared to be considerably less innovative in technology-supported instruction. Interpretation focused on the possible role of secondary level teachers' greater subject matter expertise in instructional change.

Views of project impact. ACOT teachers were reasonably satisfied with their students' progress and generally positive in their views of the values of computer use for student learning. Nevertheless, ACOT teachers appeared to feel that computers alone do not assure student engagement or meaningful learning. For example, a number of teachers were not convinced that computer use helps students stay on task, or that students cheat less when using instructional software than when working in a workbook. Similarly, at a site with comparison teachers, ACOT teachers were less likely than comparison teachers to believe that computer use helps students grasp concepts easily or encourages students to think more. ACOT teachers' critiques of the ACOT project were diverse. Many teachers remarked favorably on the value of ACOT collegial relationships and the positive impact of computer use on productive student-teacher roles in the classroom. When describing weaknesses of the project, teachers focused most often on concerns about the productivity of staff relationships (both within ACOT teams and with other teachers in school), concerns for adequate curriculum coverage, need for better coordination of teacher planning, and lack of clarity of project goals. Compared with 1989 responses, critiques in 1990 were more distinctive reflections of each site's emphasis and project progress, suggesting that emerging clarity of project goals enabled increasingly differentiated project critiques.

Views of personal impact. The project had considerable personal and professional impact on teachers, with reports of challenge and growth on the one hand and stress and demand on the other. Teachers remarked on a variety of benefits of the ACOT experience, for themselves as professionals and for their students, and appeared to be constructing new interpretations of their own and their students' abilities. Secondary teachers in particular appeared to be most actively revising their notions of students' roles in their own learning.

Effects on Parents and the Home

The ACOT parents who responded to our surveys reflected the diversity of parental backgrounds among the sites. Despite these differences in parents' education and occupation, and despite differences in ACOT project emphases, parents across sites were generally supportive of many aspects of the ACOT project and satisfied with their children's progress. Parents felt that ACOT had benefited their children's knowledge of computers, attitudes toward learning, and achievement. There were some concerns about possible trade-offs in curriculum coverage and their children's isolation from larger school community. Suggesting that parents' views of educational practice evolve as their children advance in school, parents' critiques of ACOT were more differentiated for students in upper elementary and secondary grades.

Use of the ACOT home computer. At some sites ACOT parents reported that their children spent less time in play or watching TV than did comparison parents, but these results, if valid, may have reflected the kinds of families who applied for ACOT participation. ACOT home computers were used mostly by the ACOT children themselves, although, at an urban site where few homes would otherwise have had a computer, the computers were also used often by other family members and neighbors. Interest in computer use was generally maintained over time. Most family members used the computers for a variety of functions, including published software (games, instructional drill, computer literacy training, keyboarding) and applications (e.g., word processing, graphics); for ACOT children, the range of applications in use widened at the secondary level. Family interactions involving the computer were generally positive, although parents of upper elementary and secondary level students were more apt to feel unable to participate and understand what their children were doing.

The Core Evaluation: Implications of the Results

While we had hoped that the 1990 data collection would provide definitive findings, several factors intervened to weaken our interpretations.

- The 1989-90 data collection was designed to provide us repeated measurement of progress for many ACOT and comparison students. But the unexpectedly spotty 1989-90 participation of comparison classrooms left us with empty data cells and limited understandings.

- Participating teachers were uncomfortable with certain of our core measures and requested revisions for the 1989-90 data cycle. Expressing increasing confidence in the goals for the ACOT project at their sites, teachers were concerned about potential mismatches between their goals and our assessments. We were as invested in the design of assessment alternatives as they, but we recognized that changes in our measures put at risk our planned, three-year, longitudinal evaluation design. To compromise, we reduced the testing burden on a site-by-site basis by limiting the number of grade levels tested and accepting local district test results. In addition, we selected writing prompts based on site emphases on particular genres and topics. While comparison *among* sites was never one of our objectives, comparisons over time within sites and to national norms were, and both of these were jeopardized by our design compromises.

There emerged from our results site-specific patterns in the outcome measures that begged for interpretation in the context of each site's particular focus and project thrust. But in the context of inadequate data, interpretations were more appropriately recast as questions for further study relating to school technology projects in general.

- **How does the organization and focus of a technology project affect the process of instructional change?** A number of factors of possible relevance emerged in our study. *Organizational factors* included:

  Scheduling. Longer or flexible class periods were helpful to student projects and interdisciplinary work.

  Opportunity for team collaboration and planning. The more often that teachers could work together, the better it seemed for the coherence of the program.

  School and district support. A district investment in instructional innovation or in technology appeared to ease technology-supported innovation within ACOT.

  Physical plant. Technology was more readily utilized when access was easy; innovation appeared to be facilitated when settings could be reorganized for multiple purposes and groupings.

  Differences in *project focus* included:

  Curriculum area. A selective focus seemed to foster thoughtful implementation; efforts across the curriculum appeared overly ambitious and shallow (e.g., stand-alone use of off-the-shelf software).

  Instructional goals. A focus on improving basic skills led to use of computer-based instructional software; an interest in enrichment

resulted in "problem-solving" software or student choice of software; a focus on use of computer resources for authentic engagement in productive work required applications such as word processors, graphics programs, databases, and spreadsheets.

Because these factors varied unsystematically across ACOT sites and also interacted with local contexts, firm conclusions would be premature.

- **How do teachers' knowledge and beliefs contribute to instructional uses of technology resources?** Across sites and classrooms, we found considerable variation in instructional uses of technology—variation in implementation across curricular areas, in uses for traditional (e.g., skill practice) versus nontraditional (constructive projects) student work, in choices among the many technology options available. In addition to the organizational factors already mentioned, differences among the teachers in their expertise and understandings of educational practice appeared to influence their technology implementation. For example, technology-supported innovation and experimentation appeared more likely among teachers who had not only technology skills but also—and quite critically—subject matter knowledge, curriculum knowledge (understandings of new methods of teaching a particular subject area), and competence with nontraditional pedagogy (e.g., cooperative grouping, student collaboration, long-term projects). Sound educational practice requires well-informed, solidly grounded rationale.

- **Why do students vary in their affective responses to technology?** We found site-specific patterns of impact on students' attitudes, and, where students remained in ACOT for more than one year, suggestive evidence that positive impact, for some students, was limited to the first ACOT year only. It will be important to understand how technology use can contribute to sustained, productive student engagement.

- **How should we interpret the results of our core measures?** Here we confront complex and intertwined questions regarding (a) the meaning and validity of our measures, (b) the appropriateness of our core measures to ACOT goals, and (c) the effectiveness of the ACOT projects. Beyond the limitations of our data are judgments based on values that transcend any evaluation design. Certainly the core measures do not reflect well many of ACOT's goals. But what do they reflect, and are the competences assessed by traditional measures important outcomes for any school program? If traditional achievement measures can be interpreted as valid indices of basic skills, should we be concerned that our study produced little evidence of ACOT contribution to performance? It is difficult to ignore the test results for sites where students are performing consistently and well below the national median; it is also difficult to ignore results when they show a decline (as we found at one site). These results do suggest

that ACOT programs must continue to strive to enhance certain basic skills. Nevertheless, at the same time we must move forward in the design of assessments that reflect the complex outcomes emphasized by ACOT's constructivist philosophy.

Our work with ACOT was undertaken in the early phases of what has become a paradigm shift in views on the forms and functions of assessment—away from reliance upon quantitative achievement indices representing aggregates of mastered skills and concepts, and toward performance-based measures that are intended to reflect the kinds of work valued in the workplace and community. Our questions about the fit of existing measures to ACOT teachers' objectives were to be answered by triangulating the results of different assessment approaches and by viewing the utility of any measure for the assessment of technology impact as a question for research. We knew that the design and investigation of new methods would take time. But we soon discovered that we were not alone in our quest for good measures. Our evaluation efforts corroborated the worries and resistance of teachers and administrators who were disenchanted with information produced from existing measures.

ACOT, and the challenging assessment questions it has posed, has stimulated the collaborative development of promising new measures and continues to provide a context that fosters the exploration of instructional and assessment possibilities. We next provide a brief summary of three of our research and development projects that have emerged to meet the need for appropriate and informative assessment of technology impact.

Continuing ACOT Projects

Hypermedia Measure of Knowledge Representation

Since 1989, we have been developing a new hypermedia measure of knowledge representation. Implicit in this work is a recognition that students may display competence by a range of strategies and behaviors and that present testing methodologies may seriously underrepresent the abilities of many students. It seemed clear to us that computational support might provide a means to overcoming some of the feasibility and cost problems associated with alternative measurement of complex learning outcomes such as deep understanding, explanation, and problem solving. Our hypermedia

assessment project was an outgrowth of a long-term, U.S. Department of Education-funded study of the assessment of deep understanding through the use of extended student essays. Results of prior studies (Baker, Aschbacher, Chang, Niemi, & Weinstock, 1991; Baker, Freeman, & Clayton, 1991) found that deep understanding of a subject was demonstrated by essays that were premise-driven, elaborated with text-specific and prior knowledge, and exhibited significant levels of interrelationships among facts, concepts and principles. Nonetheless, the quality of student verbal expression interacted with raters' ability to assess knowledge possessed.

To investigate whether HyperCard representations could provide a more direct measure of students' understanding, the ACOT assessment study used HyperCard as a knowledge representation tool for two studies, one involving eleventh-grade students (Baker, Niemi, & Novak, 1991) and the other eleventh and twelfth graders (Baker & Niemi, 1991). Students used specially-designed stacks to construct concept maps of history and chemistry knowledge before writing essays on given topics. Scoring systems, some of them automated, for analyzing these HyperCard products have been developed with the goal of exploring the possibility that hypertext might supplement or substitute for essay measures of deep understanding. Results suggest that one can make valid inferences about the structure of students' knowledge from HyperCard stacks and that there may be students whose knowledge is more validly represented in hypertext than in essays or standardized test items.

Problem Solving

In 1989-90, we conducted a study of students' growth in problem-solving abilities associated with the use of published "problem-solving" software. The interactive, responsive, and flexible character of computer use provides an inviting context for learning problem-solving skills. Many elementary level teachers have been attracted to instructional software packages that are designed to foster higher-level reasoning abilities such as collection, organization, and appropriate use of information when making inferences. Because there is a conflicting body of evidence in psychological and educational research on the transfer of cognitive skills from one subject matter to another, and reasonable doubt about the potential usefulness of decontextualized problem-solving practice, we felt it was important to document what kinds of

gains, if any, could be demonstrated from use of problem-solving software. We were particularly interested in whether students engaged in computer problem-solving activities would develop efficient strategies and whether these strategies would generalize to similar activities.

Based on 1988-89 field observations and teachers' descriptions of their problem-solving instruction, we designed tasks for solution both on-computer and off that were structurally and functionally similar to those practiced frequently at certain elementary ACOT sites. We administered these tasks at one elementary ACOT site during 1989-90 (fall and spring) to a sample of ACOT and comparison children who had been rated as either high or low in academic ability at one of two different grade levels. Analyses of problem-solving outcomes and strategies revealed no differences related to level of access to software. Students in the ACOT and comparison groups developed equally efficient and successful strategies, and both groups improved significantly between the fall and spring sessions. However, a subgroup of ACOT students who received special instruction in general strategies such as "Keeping track of what you know" performed better than other groups. These results provided a basis for inferring the value of problem-solving activities isolated from subject matter content as well as the role of instruction as a mediator of benefit from problem-solving software.

Portfolio Assessment of Students' Technology-Supported Compositions

When we designed our core evaluation, there existed no method of assessing the quality of ACOT students' long-term writing and multimedia projects. Therefore we initiated a collaboration with ACOT teachers to develop portfolio methods for assessing students' compositions (Baker, Gearhart, Herman, Tierney, & Whittaker, 1991). At one elementary level ACOT site, students began collecting their work in both a "working" portfolio and a smaller, student-selected "showcase" file. The showcase portfolios provided the context for an integrated set of activities: student self-assessment (prompted by sentence frames), teacher-student conferencing, informal parent-child conferencing, and parent assessment (responses to open-ended questions). There were two, somewhat disappointing, outcomes of the initial 1989-91 phase of the portfolio project—one based on documentation of

classroom practices, the second derived from an external study of the scorability of the portfolio collections.

- In the classroom, while the portfolio activities did enhance students' and teachers' interest in writing and multi-media composition, we nevertheless documented little change in teachers' assessment practices. Teachers viewed the portfolio process as a means of encouraging student investment and pride and a context for communicating with parents.

- Outside the classroom, we asked raters experienced in the assessment of traditional writing samples to judge the quality of children's portfolio collections (Gearhart, Herman, Baker, & Whittaker, 1992; Herman, Gearhart, & Baker, in preparation). Ratings were achieved with high levels of agreement and yielded the expected grade-level differences in students' competence. However, raters were able only to assign a single, holistic score; they could not assign scores for any of the analytic subscales (such as Organization or Elaboration) to the mixed collections of stories, summaries, letters, and poetry. There emerged a need to create a well-motivated match between portfolio contents and rubrics for assessment.

Therefore in 1991-92 we began the systematic development of portfolio-based assessments for children's compositions, beginning with the core genre traditionally emphasized at the elementary level. By working with grade-level teams (K-1, 2-3, and 4-6) as well as a schoolwide steering committee, our goals are to develop agreement at the school level on standards that constitute writing competence and methods to track students' progress toward those standards (Dietel, Gearhart, & Herman, 1992; Gearhart, Herman, Baker, Wolf, & Whittaker, 1992; Gearhart, Herman, Wolf, & Baker, 1992).

A key component of the curriculum and assessment framework—*Writing What You Read* (Wolf & Gearhart, in preparation)—is the analysis of *exemplars of excellence*. For narrative, the exemplars are drawn from children's literature appropriate to each grade-level team. Teachers are learning to analyze literature using a technical language that enables more substantive and more explicit advice for guiding growth in young writers. Children are organizing their portfolios by genre (e.g., stories, reports, persuasive letters, poetry), and teachers' and students' evaluations of writing progress are supported by assessment forms that represent the components of good writing in each genre and by developmental rubrics that reflect dimensions or continua of change rather than discrete skills.

Summary

Our evaluation studies of five Apple Classrooms of Tomorrow sites have documented the challenges of assessing the effects of technology access on teaching and learning. In the context of evolving ACOT goals and site-specific project strategies, our strategy was to employ a coordinated set of standard measures as we developed additional tools to capture emerging modes of instruction and unanticipated outcomes.

While not conclusive, our work uncovered patterns of effects that suggest new lines of inquiry. We close with two examples. First, positive findings that ACOT may have had *selective* positive impact on writing and on student attitudes, particularly as projects matured, made evident the need to examine the impact of *specific* instructional uses of technology on *specific* outcomes. Second, findings that changes in practice were consistent with ACOT goals, such as increased use of cooperative groups and decreased "stand-up" directive teaching, did not reveal the *particular role of technology* in these changes. These emergent questions affirm the critical importance of measures sensitive to particular technology project designs and contexts.

References

Baker, E. L. (1987, September). *Time to write: Report of the US-IEA Study of Written Composition.* Invited presentation to the IEA General Assembly, Teachers College, New York.

Baker, E. L., & Aschbacher, P., Chang, S. C., Niemi, D., & Weinstock, M. (1991, April). *Validating measures of deep understanding of history scored by teachers.* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Baker, E. L., Freeman, M., & Clayton, S. (1991). Cognitive assessment of history for large-scale testing. In M. C. Wittrock & E. L. Baker (Eds.), *Testing and cognition*. Englewood Cliffs, NJ: Prentice-Hall.

Baker, E. L., Gearhart, M., & Herman, J. L. (1990). *The Apple Classrooms of Tomorrow: 1989 Evaluation study* (Report to Apple Computer, Inc.). Los Angeles: University of California, Center for the Study of Evaluation/ Center for Technology Assessment.

Baker, E. L., Gearhart, M., & Herman, J. L. (1992). *The Apple Classrooms of Tomorrow: 1990 Evaluation study* (Report to Apple Computer, Inc.). Los Angeles: University of California, Center for the Study of Evaluation/ Center for Technology Assessment.

Baker, E. L., Gearhart, M., Herman, J. L., Tierney, R., & Whittaker, A. K. (1991). Stevens Creek portfolio project: Writing assessment in the technology classroom. *Portfolio News, 2*(3), 7-9.

Baker, E. L., Herman, J. L., & Gearhart, M. (1988). *The Apple Classrooms of Tomorrow: 1988 Evaluation study* (Report to Apple Computer, Inc.). Los Angeles: University of California, Center for the Study of Evaluation/ Center for Technology Assessment.

Baker, E. L., Herman, J. L., & Gearhart, M. (1989, April). *The ACOT report card: Effects on complex performance and attitude.* Presentation at the 1989 annual meeting of the American Educational Research Association, San Francisco.

Baker, E. L., & Niemi, R. (1991, April). *Assessing deep understanding of history and science through hypertext.* Paper presented at the annual meeting of the America Educational Research Association, Chicago.

Baker, E. L, Niemi, J., & Novak, J. (1991, July). *Hypertext as a strategy for teaching and assessing knowledge representation.* Invited paper presentation at the NATO Advanced Research Workshop, Enschede, The Netherlands.

Dietel, R., Gearhart, M., & Herman, J. L. (1992). *Portfolio assessment and high technology* (Video). Los Angeles: University of California, Center for the Study of Evaluation

Gearhart, M., Herman, J. L., Baker, E. L., Novak, J. R., & Whittaker, A. K. (in press). A new mirror for the classroom: A technology-based tool for documenting the impact of technology on instruction. In E. L. Baker & H. F. O'Neil, Jr. (Eds.), *Technology assessment. Vol. 1: Technology and training.* Hillsdale NJ: Erlbaum.

Gearhart, M., Herman, J. L., Baker, E. L., & Whittaker, A. K. (1992). *Writing portfolios at the elementary level: A study of methods for writing assessment* (CSE Tech. Rep. No. 337). Los Angeles: University of California, Center for the Study of Evaluation.

Gearhart, M., Herman, J. L., Baker, E. L., Wolf, S. A., & Whittaker, A. K. (1992, April). *Portfolios: An approach to the assessment of technology-supported composition.* Paper presented for the symposium entitled "Understanding technology impact in schools: New approaches to student assessment" at the 1992 annual meeting of the American Educational Research Association, San Francisco.

Gearhart, M., Herman, J. L., Whittaker, A. K., & Novak, J. R. (1991, April). *The effects of high access on instruction.* Paper presented for the symposium entitled "The impact of technology on classroom activities and interactions: Restructuring instruction?" at the 1991 annual meeting of the American Educational Research Association, Chicago.

Gearhart, M., Herman, J. L., Wolf, S. A., & Baker, E. L. (1992, April). *Writing portfolios at the elementary level: A study of methods for writing assessment.* Paper presented for the symposium entitled "Perspectives on the assessment value of portfolios: How robust are they?" at the 1992 annual meeting of the American Educational Research Association, San Francisco.

Herman, J. L, Gearhart, M., & Baker, E. L. (in preparation). *Assessing writing portfolios: Issues in the validity and meaning of scores.* Manuscript, University of California, Los Angeles, Center for the Study of Evaluation.

Wolf, S. A., & Gearhart, M. (in preparation). *Writing What You Read: Guiding growth in young writers.* Manuscript, University of California, Los Angeles, Center for Research on Evaluation, Standards, and Student Testing.